# Ghaleb M. Abdulla

**Center for Applied Scientific Computing, L-560**
**Lawrence Livermore National Laboratory**
**Livermore, CA 94551**

**Phone: (925) 423-5947**
**Fax: (925) 422-6287**
**E-mail:abdulla1@llnl.gov**


## Research Interests

- Large Scale Data Management and Data Mining

## Education

| | |
|---|---|
| Ph.D. | Computer Science, Virginia Tech, Blacksburg, VA, 1998 |
| | Thesis: "Analysis and Modeling of World Wide Web Traffic" |
| | Advisor: Dr. Edward A. Fox |
| M.S. | Computer Science, Virginia Tech, Blacksburg, VA, 1993 |
| | Thesis: "A Tool for Enhancing and Cropping Images" |
| | Advisor: Dr. Edward A. Fox |
| B.S. | Electrical and Computer Engineering, Yarmouk University, Irbid, Jordan, 1984 |

## Professional Experience

| | |
|---|---|
| 12/03-present | Technical Leader/Senior Scientist, Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA |
| 09/07-12/2007 | Visiting faculty, Electrical and Computer Engineering Department, University of the Pacific |
| 10/00–12/03 | Computer Scientist, Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA |
| 6/98–9/00 | Information Technology Specialist, The Dow Chemical Company, Business Intelligence Center, Freeport, TX |
| 1/91–5/98 | Research and Teaching Assistant, Department of Computer Science, Virginia Tech, Blacksburg, VA |
| 8/88–8/90 | Senior Software Engineer, Arab Fund for Economical and Social Development, Kuwait |
| 2/84–8/88 | Senior Systems Engineer, Kuwait |

## Skills

Python, R, Cloudera Hadoop, Excel, mpich, C, C++ , java

## Honors and Organizations

- Sigma Xi, The Scientific Research Society
- Upsilon pi upsilon, Honor Society of the Computer Sciences
- ACM/IEEE
- Emerging Leadership award (LLNL)
- Service Award (LLNL)
- Spot award (LLNL)
- Service award (Dow Chemical Company)

## Current Projects

**1- Understanding the stochastic nature of laser-induced damage**

Duration: 09/2008-

- Understand and predict laser damage growth

Publications:

1- Delmar B. Davis, Hazeline U. Asuncion, Ghaleb Abdulla "Experiment Explorer: Lightweight Provenance Search over Metadata", TaPP 2012 : 4th USENIX Workshop on the Theory and Practice of Provenance
2- Zhi M. Liao, G. M. Abdulla, R. A. Negres, D. A. Cross, and C. W. Carr, "Predictive Modeling Techniques for nanosecond-laser damage growth in fused silica optics", to appear Opt. Express 2012
3- Raluca A. Negres, Ghaleb M. Abdulla, David A. Cross, Zhi M. Liao, and Christopher W. Carr, "Probability of growth of small damage sites on the exit surface of fused silica optics", Opt. Express 20, 13030-13039 (2012)
4- Ghaleb Abdulla, Abdul Awwal, Kirk Borne, Tin Kam Ho, and W. Thomas Vestrand, "Practical data mining and machine learning for optics applications: introduction to the feature issue", Applied Optics Journal, August 2011
5- Raluca A. Negres, Z. M. Liao, G. M. Abdulla, D. A. Cross, M. A. Norton, and C. W. Carr, "Exploration of the multi-parameter space of ns-laser damage growth in fused silica optics", Applied Physics Journal, August 2011
6- Ghaleb M. Abdulla, Laura M. Kegelmeyer, Zhi M. Liao, Christopher W. Carr, Use of machine learning algorithms and data mining techniques to identify and remove noisy data, SPIE XLII Annual Symposium on Optical Materials for High Power Lasers, 26-29 September 2010 Proceedings of SPIE Vol. 7842 (2010)
7- A. Carr, L Kegelmeyer, Z. M. Liao, G. Abdulla, D. Cross, W. P. Kegelmeyer, F. Ravizza, and C. W. Carr, "Defect Classification Using Machine Learning" in Laser-Induced Damage in Optical Materials: 2008, G. J. Exarhos, A. H. Guenther, K. L. Lewis, D. Ristav, M. J. Soileau, and C. J. Stolz, eds., Proc. SPIE 7132 (2008)

**2- Computational Challenges for the smart Grid implementation**

Duration: 10/2010-

- Investigate the role of of HPC and data intensive computing in the smart grid
- Define requirements and strategy to create an energy informatics center

During the summer of 2011, I worked at the CAISO site for three months (2 days a week) where I gave feedback on the smart grid road map plans report, identified areas of collaboration especially in the area of renewable integration, and got familiar with their market and grid models,

Publications:

1- Sean Crimmins, Yinka Osoba, Doug Walker, Kevin Colmer and Ghaleb Abdulla, "Smart Grid Computational and Modeling Challenges", Computational Challenges 2011: Energy Resource Modeling, August 2011, Poster

## Previous Projects

### Predictive Knowledge systems/ CoPI

Duration: 05/2006-09/2008

Funding level: 3M/year

- Investigated the scalability of graph queries using a traditional RDBMS system implementation and performed scalability tests using a suite of queries that we identified to be most relevant for our application.

- Investigated scalability of spatial matching for a large set of objects using different database and hardware configuration including Netezza.

Publications:

1. Vijay S. Kumar, Tahsin Kurc, Ghaleb Abdulla, Scott R. Kohn, Joel Saltz and Celeste Matarazzo **"**Architectural Implications for Spatial Object Association Algorithms**",** IEEE International Parallel & Distributed Processing Symposium, 2009
2. Benchmark queries for large-scale semantic graph applications, Justin Levandoski, Ghaleb Abdulla, Ian Kaplan, and Scott Kohn, Technical Report
3. Nikolaev S., Ghaleb Abdulla, Robb Matzke, "Evaluation of Potential Large Synoptic Survey Telescope Spatial Indexing Strategies", 2006, Technical Report, UCRL-TR-225827.

**Progressive algorithms and novel architectures for real-time analysis of streaming data/PI**

Duration: 05/2004- 01/2007

Funding level: 225K/year outside funding + 400K inside research funding

Lead a team of 3 researchers and 1 graduate student to investigate a software framework and to build a prototype that supports data-centric analysis for large and streaming image data. We also examined implementation issues of such a framework onto cluster computing and high performance file systems. In the process we worked on developing a prototype of a large-scale database for astronomical data using precursor data sets. We investigated bulk loading, efficient indexing, and spatial and temporal querying of the database. My rule as a PI was to propose the basic research ideas, lead the team, and initiate external collaborations with experts in the field such as Jim Gray from Microsoft, Michael Franklin from Berkeley and Jacek Becla from SLAC.

Related publications and presentations:

1. Liu, David T., Abdulla, Ghaleb M., Franklin, Michael J., Garlick, Jim, Miller, Marcus M., Data-Preservation in Scientific Workflow Middleware, 18th International Conference on Statistical and Scientific Databases, 2006
2. Becla J., Andrew Hanushevsky, Sergei Nikolaev, Ghaleb Abdulla, Alex Szalay, Maria Nieto-Santisteban, Ani Thakar, Jim Gray, "Designing a multi-petabyte database for LSST" (SPIE 2006)
3. Becla J., S. Nikolaev, G. Abdulla, A.S. Szalay, M. Nieto-Santisteban, A. Thakar, T. Axelrod, J. Gray, R. Pike, W. Rosing "LSST Data Access Overview" (AAS 2006)
4. Liu, D., M. Franklin, J. Garlick, G. M. Abdulla, M. Miller, "Scaling Up Data Provenance and Smart Recomputation for Scientific Workflows", Proceeding of *SC05*, Seattle, Washington, November 12-18, 2005
5. Abdulla, G., D. Liu, J. Garlick, S. Nikolaev, M. Miller, M. Franklin, K. Cook, J. Brase, "A Prototype of the Future LSST Data Pipeline" Astronomical Data Analysis Software & Systems XIV October 24 - 27, 2004 Pasadena California
6. Nikolaev S., M. E. Huber, K. H. Cook, G. Abdulla, J. Brase, "Conceptual Design of a Prototype LSST Database", 2006, Technical Report, UCRL-TR-207256
7. Nikolaev S., Ghaleb Abdulla, Robb Matzke, "Evaluation of Potential Large Synoptic Survey Telescope Spatial Indexing Strategies", 2006, Technical Report, UCRL-TR-225827
8. Abdulla G. "Analysis of SDSS SQL server log files", 2006, Technical Report, UCRL-MI-215756-DRAFT

**Analysis of scientific simulation data:**

The aim of this research was to reduce the size of the data sets being analyzed by using statistical models, allowing a broad range of user defined queries, and determining appropriate ways to store data for efficient retrieval. In this research I worked on the design and implementation of the

query engine for the datafoundry project which aims at helping scientists visualize their data by supporting ad hoc queries. I did the analysis of the disk access and helped design an optimized version of the node implementation. I designed and helped implement the parallel version of the system using MPI. The implementation was tested on MCR which at that time was the fastest supercomputer in the world. The research was funded under the Advanced Scientific Computing Research program (ASCI).

Publications and presentations:

1. Abdulla, G., W. Arrighi, T. Critchlow, "Scientific Simulation Data as Data streams", SIGMOD Record, Vol. 33 Num 1, March 2004, PP. 89-94.
2. T. Eliassi-Rad, C. Baldwin, G. Abdulla, and T. Critchlow. Statistical Modeling of Large-Scale Scientific Simulation Data. *New Generation of Data Mining Applications*, Eds: J. Zurada and M. Kantardzic, IEEE Press/Wiley Publishers, 2005.
3. B. Lee, T. Critchlow, G. Abdulla, C. Baldwin, R. Kamimura, R. Musick, R. Snapp, and N. Tang, "The Framework for Approximate Queries on Simulation Data," *Journal of Information Sciences*, Elsevier Sciences, Vol. 157, Issues 1-4, December 2003, pp. 3-20.
4. Baldwin, C., G. Abdulla, and T. Critchlow, " Multi-Resolution Modeling of Large scale Scientific Simulation Data," *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, New Orleans, LA, November 3-8 2003
5. Kamimura, R., G. Abdulla, C. Baldwin, T. Critchlow, B. Lee, I. Lozares, R. Musick, and N. Tang, "Use of Numerical Models as Data Proxies for Approximate Ad-Hoc Query Processing," *Proceedings of the 7th International Conference on Computer Science and Informatics*, Cary, North Carolina, U.S.A., September 26 – 30, 2003.
6. Baldwin, C., T. Eliassi-Rad, G. Abdulla, and T. Critchlow, "*The Evolution of a Hierarchical Partitioning Algorithm for Large-Scale Scientific Data: Three Steps of Increasing Complexity*," Fifteenth Int. Conf. on Scientific and Statistical Data Base Management (SSDBM 2003), Cambridge, MA, July 2003.
7. Tina Eliassi-Rad, Terence Critchlow, and Ghaleb Abdulla, "Statistical Modeling of Large-Scale Simulation Data," Eighth ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD 2002), Edmonton, Alberta, Canada, July 2002.
8. Abdulla, G., C. Baldwin, T. Critchlow, R. Kamimura, I. Lozares, R. Musick, N. Tang, B. Lee, R. Snapp, "Approximate Ad-Hoc Query Engine for Simulation Data," *Proc. Joint Conference on Digital Libraries JCDL-01*. June 2001.

**Business intelligence, data integration and knowledge management:**

At Dow chemical I worked as an information technology specialist in the business intelligence center. We worked on creating a digital library for the company's internal technical reports. We also looked into integrating the internal information with publicly available data in order to make informed decisions about business directions. I also worked on a data mining project to mine the internal technical report collection over the last 20 years. We looked at the trends in the research and how it changes overtime.

Related publications:

1. Rogers, A., G. Abdulla, K. Rapp, G. Hoffman, C. J. Scott, and V. Tanner, "Dow Technology Enhances CRI Reporting Process," *The Dow Chemical Company Science and Technology Meeting*, Midland, MI, May 24, 2000.
2. McNamee, G, G. Abdulla, and C. Mott, "Data Mining/Visualization in Polypropylene R&D," *The Dow Chemical Company Science and Technology Meeting,"* Gulf Coast

Scientists, Brazosport Center for Arts and Sciences, Lake Jackson, TX, September 30, 1999.

**WWW traffic modeling and characterization**

We developed a modeling approach that considers Web traffic characteristics such as self-similarity and long-range dependency. We developed an algorithm to characterize users' sessions and high-level Web traffic model suitable for sensitivity analysis. We studied the temporal and spatial locality of reference within the examined user communities. We utilized the findings to recommend caching algorithms since it can be an effective tool to help reduce network traffic and to help solve the web scalability problem. We recommend utilizing our findings to promote a smart distribution or push model to cache documents when there is likelihood of repeat accesses. My rule was to design and write software to collect network data and parse it into a format similar to a web log file. I also wrote MATLAB, Perl, and S-Plus data analysis functions. I used Expert-fit to uncover the underlying statistical distributions of the various types of collected data.
Related publications:

1. Abdulla, G., L. Binzhang, and E.A. Fox, "Searching the WWW: Implications from Studying Different User Behavior," *ProcNet98 Conference*, Orlando, FL, November 7-12, 1998. http://www.aace.org/conf/webnet, pp. 1-6.
2. Binzhang, L., G. Abdulla, T. Johnson, and E.A. Fox, " Web Response Time and Proxy Caching," *Proc. Webnet98 Conference*, Orlando, FL, November 7-12, 1998. http://www.aace.org/conf/webnet, pp. 591-596, recognized as "top paper."
3. Habib, M.D., G. Abdulla, and E. Fox, "Web Traffic Characterization with Time Zones: Seeking Outside Events that Affect the Traffic to a Distance Learning Server," *Proc. international Symposium on Audo, Video, Image Processing, and Intelligent Application* (ISAVIIA-98), Germany, August 17-21, 1998.
4. Abdulla, G., M. Abrams, and E.A. Fox, "Shared User Behavior on the World Wide Web," *Proc. of WebNet97*, Toronto, Canada, October 1997.
5. Abdulla, G., W. Heagy, and E.A. Fox, "Quantitative Analysis and Visualization Regarding Interactive Learning with a Digital Library in Computer Science," poster in *Second ACM International Conference on Digital Libraries*, Philadelphia, PA, July 1997.
6. Williams, S; Abrams, M; Strandridge, C R; Abdulla, G; Fox, E A "Removal policies in network caches for world-wide web documents", Computer Communication Review. Vol. 26, no. 4, pp. 293-305. 1996, reprinted from *ACM SIGCOMM'96 Conference*, Stanford University, CA, August 1996, pp. 293-305.
7. Abrams, M., C.R. Standridge, G. Abdulla, S. Williams, and E.A. Fox, "Caching Proxies: Limitations and Potentials," *World Wide Web* **1** (1995), pp. 119-133. Reprinted from *proc. 4th International World Wide Web Conference*, Boston, December 1995.

## Professional Activities

- Reviewer for WWW Journal, IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Software Engineering, Journal of Software Practice and Experience, the journal of systems and software, and IEEE Internet Computing Journal
- The First ACM IEEE Joint Conference on Digital libraries 2001, Student Volunteer Chair
- Joint Conference on Digital Libraries 2002, 2003, 2004, 2005, 2006 Program Committee
- Program committee SIGIR 2007, 2008 and CIKM 2007, ACM GIS 2008, 2009
- NSF Medium ITR Panelist, NSF Information & Knowledge Management panelist

- DOE MAPD and early career award panelist
- UC Riverside computer science and electrical engineering department advisory board member
- San Jose state university computer science advisory board member

## Professional Training

### Technical

- Multivariate Statistical Methods for the Analysis, Monitoring, and Optimization of Processes
- Programming with MATLAB 5

### Management

- UCLA extensions technical management program
  - Influencing without authority
  - Strategic project thinking: Designing action plans to achieve critical objectives
  - Leadership and management skills for intensely competitive and rapidly changing times
  - Entrepreneurship and risk taking in technology based organizations
- Emerging Leaders program
  - The 7 Habits of highly effective people
  - The 4 Roles of Leadership
  - How to handle difficult people
  - Organizing Genius: The Secrets of Creative Collaboration (Warren Bennis)

## Other Publications and Presentations

### Book Chapter

1. Patel, S., G. Abdulla, M. Abrams, and E. Fox, "NMFS: Network Multimedia File System Protocol," in P. Venkat Rangan, *Network and Operating System Support for Digital Audio and Video*, Lecture Notes in Computer Science 712 (Springer-Verlag, Berlin Heidelberg, 1993) ISBN 3-540-57183-3 and 0-387-57183-3, pp. 328-333. Reprinted from conference proceedings of *3rd International Workshop on Network and Operating System Support for Digital Audio and Video*, La Jolla, CA, Nov. 12-13, 1992.

#### Conference Papers/Poster

1. Ghaleb Abdulla, Eric Estephan, Sergei Nikolaiv, Tina Eliassi-Rad, Terence Critchlow, "Similarity Queries Over Intensive Data, Challenges Approaches and Results". submitted for publication.
2. Abrams M., E. A. Fox, G. Abdulla, S. Williams, *"Experience in Network Delivery of Computer Science Courseware*," transparencies from presentation at the 2nd Annual SUCCEED Conference, N.C. State, March 1995.
3. Abrams, M., S. Williams, G. Abdulla, S. Patel, R. Ribler, and E.A. Fox, "Multimedia Traffic Analysis Using Chitra95," *Proc. ACM Multimedia '95*, San Francisco, November 1995, pp. 267-276.
4. Fox, E. and G. Abdulla, "Digital Delivery for a Digital Library in Computer Science," *High-Speed Networking and Multimedia Computing Workshop*, IS&T/SPIE Symposium on Electronic Imaging Science and Technology, San Jose, CA, February 6-10, 1994.

**Invited Presentations**

1. Abdulla, Ghaleb, “Data Intensive Applications, Challenges and Potentials”, University of Minnesota, February 13, 2006
2. Abdulla, Ghaleb, “A middleware to support analysis of astronomy data sets”, ADASS04 meeting, Pasadena, CA, October 24-28
3. Abdulla, Ghaleb, "Query Pattern Analysis," *SciDAC Hands on Mtg.*, invited talk, Gatlinburg, TN, March 26-27, 2002. Also available as Lawrence Livermore National Laboratory technical report UCRL-PRES-147651, April 2002.
4. Abdulla, Ghaleb, and Tina Eliassi-Rad, "Query Pattern Analysis," *SciDAC All Hands Mtg.*, Also available as Lawrence Livermore National Laboratory technical report UCRL-PRES-149909, September 2002.
5. Abdulla, Ghaleb, “Querying Tera-Scale Scientific Simulation Data,” University of Central Florida, invited talk, Orlando, Florida, December 28, 2001.
6. Abdulla, G., “Approximate Ad-hoc Query Engine for Simulation Data,” University of Vermont, September 10, 2001.
7. Abdulla, G., “Querying Scientific Data,” SciDAC kickoff meeting, Lawrence Berkeley Lab., July 10, 2001.
8. Abulla, G., “Approximate Ad-hoc Query Engine for Simulation Data,” Virginia Polytechnic Institute and State University, June 29, 2001.
9. Abdulla, G., “WWW Proxy Traffic Modeling,” *Seminar Series*, Computer Science Department, University of Houston, TX, March 2, 2000.
10. Abdulla, G., “Web Traffic Modeling,” *The W3C Web Characterization Workshop*, Cambridge, MA, November 5, 1998.
11. Fox, E.A., G. Abdulla, and N. Kipp, “ DL Metrics: Web Characterization and 4S,” *Digital Library Metrics Workshop*, June 27, 1998.

**Technical Reports**

1. Abdulla G. “Analysis of SDSS SQL server log files”, 2006, Technical Report, UCRL-MI-215756-DRAFT
2. Nikolaev S., Ghaleb Abdulla, Robb Matzke, "Evaluation of Potential Large Synoptic Survey Telescope Spatial Indexing Strategies", 2006, Technical Report, UCRL-TR-225827
3. Nikolaev S., M. E. Huber, K. H. Cook, G. Abdulla, J. Brase, "Conceptual Design of a Prototype LSST Database", 2006, Technical Report, UCRL-TR-207256
4. Chuck Baldwin, Ghaleb Abdulla, William Arrighi, Susan Hazlett, Megan Thomas, Tina Eliassi-Rad, and Terence Critchlow, “Finding the Needle in the Haystack: Performing Queries and Analysis on Large Scale Scientific Simulation Data”, *UCRL-CONF-203894*, Lawrence Livermore National Lab
5. Abdulla, G., A.H. Nayfeh, and E.A. Fox, “Modeling Correlated proxy Web Traffic Using Fourier Analysis,” technical report *TR-97-19*, Computer Science Department, Virginia Tech, November 1997
6. Abdulla, G., L. Binzhang, R. Saad, and E.A. Fox, “Characterizing WWW Queries,” technical report *TR-97-04*, Computer Science Department, Virginia Tech, March 1997.

7. Abdulla, G., M. Abrams, E.A. Fox, and Stephen Williams, “WWW Proxy Traffic Characterization with Application to Caching,” technical report *TR-97-03*, Computer Science Department, Virginia Tech, March 1997.
8. Abdulla, G., M. Abrams, and E.A. Fox, “Scaling the WWW,” technical report *TR-96-06*, Computer Science Department, Virginia Tech, March 1996.